

1 CGACCTGGCC GCCGGCCGCT CCTCCGCGCG CTGTTCCGCA CTTGCTGCCC
51 TCGCCCGGCC CGGAGCGCCG CTGCCATGCG GCTGGCGCTG CTCTGGGCCC
101 TGGGGCTCCT GGGCGCGGGC AGCCCTCTGC CTTCTGGGCC GCTCCCAAAT
151 ATAGCCCTGC TGTCGATTCC CTCAGTACTG TCTTGGGGTG TCCTGGGACC
201 TGCAGGTGGC ACTGAGGAGC AGCAGGCAGA GTCAGAGAAG GCCCCGAGGG
251 AGCCCTTGGA GCCCCAGGTC CTTCAGGACG ATCTCCCAAT TAGCCTCAAA
301 AAGGTGCTTC AGACCAGTCT GCCTGAGCCC CTGAGGATCA AGTTGGAGCT
351 GGACGGTGAC AGTCATATCC TGGAGCTGCT ACAGAATAGG GAGTTGGTCC
401 CAGGCCGCCC AACCCTGGTG TGGTACCAGC CCGATGGCAC TCGGGTGGTC
451 AGTGAGGGAC ACACTTTGGA GAACTGCTGC TACCAGGGAA GAGTGCGGGG
501 ATATGCAGGC TCCTGGGTGT CCATCTGCAC CTGCTCTGGG CTCAGAGGCT
551 TGGTGGTCCT GACCCAGAG AGAAGCTATA CCCTGGAGCA GGGGCCTGGG
601 GACCTTCAGG GTCCTCCAT TATTTCCGCA ATCCAAGATC TCCACCTGCC
651 AGGCCACACC TGTGCCCTGA GCTGGCGGGA ATCTGTACAC ACTCAGACGC
701 CACCAGAGCA CCCCCTGGGA CAGCGCCACA TTCGCCGGAG GCGGGATGTG
751 GTAACAGAGA CCAAGACTGT GGAGTTGGTG ATTGTGGCTG ATCACTCGGA
801 GGCCAGAAA TACCGGGACT TCCAGCACCT GCTAAACCGC ACACTGGAAG
851 TGGCCCTCTT GCTGGACACA TTCTTCCGGC CCCTGAATGT ACGAGTGGCA
901 CTAGTGGGCC TGGAGGCCTG GACCCAGCGT GACCTGGTGG AGATCAGCCC
951 AAACCCAGCT GTCACCCTCG AAAAATTCTT CCACTGGCGC AGGGCACATT
1001 TGCTGCCTCG ATTGCCCCAT GACAGTGCCC AGCTGGTGAC TGGTACTTCA
1051 TTCTCTGGGC CTACGGTGGG CATGGCCATT CAGAACTCCA TCTGTTCTCC
1101 TGACTTCTCA GGAGGTGTGA ACATGGACCA CTCCACCAGC ATCCTGGGAG
1151 TCGCCTCCTC CATAGCCCAT GAGTTGGGCC ACAGCCTGGG CCTGGACCAT
1201 GATTTGCCTG GGAATAGCTG CCCCTGTCCA GGTCCAGCCC CAGCCAAGAC
1251 CTGCATCATG GAGGCCTCCA CAGACTTCCT ACCAGGCCTG AACTTCAGCA
1301 ACTGCAGCCG ACGGGCCCTG GAGAAAGCCC TCCTGGATGG AATGGGCAGC
1351 TGCCTCTTCG AACGGCTGCC TAGCCTACCC CCTATGGCTG CTTTCTGCGG
1401 AAATATGTTT GTGGAGCCGG GCGAGCAGTG TGA CTGTGGC TTCCTGGATG
1451 ACTGCGTCGA TCCCTGCTGT GATTCTTTGA CCTGCCAGCT GAGGCCAGGT
1501 GCACAGTGTG CATCTGACGG ACCCTGTTGT CAAAATTGCC AGCTGCGCCC
1551 GTCTGGCTGG CAGTGTGCTC CTACCAGAGG GGATTGTGAC TTGCCTGAAT
1601 TCTGCCCAGG AGACAGCTCC CAGTGTCCCC CTGATGTCAG CCTAGGGGAT
1651 GGCGAGCCCT GCGCTGGCGG GCAAGCTGTG TGCATGCACG GGC GTTGTGC
1701 CTCCTATGCC CAGCAGTGCC AGTCACTTTG GGGACCTGGA GCCCAGCCCC
1751 CTGCGCCACT TTGCCTCCAG ACAGCTAATA CTCGGGGAAA TGCTTTTGGG
1801 AGCTGTGGGC GCAACCCAG TGGCAGTTAT GTGTCCTGCA CCCCTAGAGA
1851 TGCCATTTGT GGGCAGCTCC AGTGCCAGAC AGGTAGGACC CAGCCTCTGC
1901 TGGGCTCCAT CCGGGATCTA CTCTGGGAGA CAATAGATGT GAATGGGACT
1951 GAGCTGAACT GCAGCTGGGT GCACCTGGAC CTGGGCAGTG ATGTGGCCCA
2001 GCCCCTCCTG ACTCTGCCTG GCACAGCCTG TGGCCCTGGC CTGGTGTGTA
2051 TAGACCATCG ATGCCAGCGT GTGGATCTCC TGGGGGCACA GGAATGTCGA
2101 AGCAAATGCC ATGGACATGG GGTCTGTGAC AGCAACAGGC ACTGCTACTG
2151 TGAGGAGGGC TGGGCACCCC CTGACTGCAC CACTCAGCTC AAAGCAACCA
2201 GCTCCCTGAC CACAGGGCTG CTCCTCAGCC TCCTGGTCTT ATTGGTCTG
2251 GTGATGCTTG GTGCCAGCTA CTGGTACCGT GCCCGCCTGC ACCAGCGACT

FIGURE 1A

2301 CTGCCAGCTC AAGGGACCCA CCTGCCAGTA CAGGGCAGCC CAATCTGGTC
2351 CCTCTGAACG GCCAGGACCT CCGCAGAGGG CCCTGCTGGC ACGAGGCACT
2401 AAGGCTAGTG CTCTCAGCTT CCCGCCCCC CCTTCCAGGC CGCTGCCGCC
2451 TGACCCTGTG TCCAAGAGAC TCCAGTCTCA GGGGCCAGCC AAGCCCCCAC
2501 CCCCAAGGAA GCCACTGCCT GCCGACCCCC AGGGCCGGTG CCCATCGGGT
2551 GACCTGCCCCG GCCCAGGGGC TGGAAATCCCG CCCCTAGTGG TACCCTCCAG
2601 ACCAGCGCCA CCGCCTCCGA CAGTGTCTC GCTCTACCTC TGACCTCTCC
2651 GGAGGTTCCG CTGCCTCCAA GCCGGACTTA GGGCTTCAAG AGGCGGGCGT
2701 GCCCTCTGGA GTCCCCTACC ATGACTGAAG GCGCCAGAGA CTGGCGGTGT
2751 CTTAAGACTC CGGGCACCGC CACGCGCTGT CAAGCAACAC TCTGCGGACC
2801 TGCCGGCGTA GTTGCAGCGG GGGCTTGGGG AGGGGCTGGG GGTTGGACGG
2851 GATTGAGGAA GGTCCGCACA GCCTGTCTCT GCTCAGTTGC AATAAACGTG
2901 ACATCTTGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2951 AAAAAAAAAA AAAAAAAA
(SEQ ID NO: 1)

FEATURES:

5'UTR: 1 - 75
Start Codon: 76
Stop Codon: 2641
3'UTR: 2644

Homologous proteins:

Top 10 BLAST Hits:

Sequences producing significant alignments:	Score	E
Value	(bits)	
CRA 335001098640323 /altid=gi 7451525 /def=pir G02390 disinteg...	1714	0.0
CRA 335001098639998 /altid=gi 11497002 /def=ref NP_003806.2 a ...	1698	0.0
CRA 1000682348196 /altid=gi 9945328 /def=ref NP_064704.1 a dis...	1377	0.0
CRA 18000005154484 /altid=gi 6752962 /def=ref NP_033744.1 a di...	1351	0.0
CRA 1000737073449 /altid=gi 6682839 /def=dbj BAA88903.1 (AB022...	1319	0.0
CRA 157000140328366 /altid=gi 12720142 /def=ref XP_010635.1 a ...	970	0.0
CRA 18000005119563 /altid=gi 4501905 /def=ref NP_003465.1 a di...	539	e-152
CRA 98000043629034 /altid=gi 13027660 /def=gb AAC08702.2 (AF02...	539	e-152
CRA 18000005009258 /altid=gi 6680640 /def=ref NP_031426.1 a di...	538	e-151
CRA 98000043606871 /altid=gi 12802370 /def=gb AAK07852.1 AF3113...	517	e-145

FIGURE 1B

EST:

Sequences producing significant alignments:
Value

Score E
(bits)

gi 12777372 /dataset=dbest /taxon=960...	1750	0.0
gi 10205626 /dataset=dbest /taxon=96...	1364	0.0
gi 10746030 /dataset=dbest /taxon=96...	1352	0.0
gi 12758166 /dataset=dbest /taxon=960...	1334	0.0
gi 13130161 /dataset=dbest /taxon=960...	1306	0.0
gi 11003698 /dataset=dbest /taxon=96...	1298	0.0
gi 12763891 /dataset=dbest /taxon=960...	1281	0.0
gi 9124688 /dataset=dbest /taxon=9606...	1211	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

gi|12777372 placenta
gi|10205626 lung
gi|10746030 ovary
gi|12758166 colon
gi|13130161 kidney
gi|11003698 thyroid gland
gi|12763891 prostate
gi|9124688 eye

Tissue expression:
leucocyte

1 MRLALLWALG LLGAGSPLPS WPLPNIALLS IPSVLSWGLV GPAGGTEEQQ
51 AESEKAPREP LEPQVLQDDL PISLKKVLQT SLPEPLRIKL ELDGDSHILE
101 LLQNRELVPQ RPTLVWYQPD GTRVVSEGHT LENCCYQGRV RGYAGSWVSI
151 CTCSGLRGLV VLTPEPSYTL EQGPGDLQGP PIISRIQDLH LPGHTCALS
201 RESVHTQTPP EHPLGQRHIR RRRDVTETK TVELVIVADH SEAQKYRDFQ
251 HLLNRTLEVA LLLDTFFRPL NVRVALVLE AWTQRDLVEI SPNPAVTLEN
301 FLHWRRRAHLL PRLPHDSAQL VTGTSFSGPT VGMAIQNSIC SPDFSGGVNM
351 DHSTSILGVA SSIAHELGH LGLDHDLPGN SCPCPGPAPA KTCIMEASTD
401 FLPGLNFSNC SRRALEKALL DGMGSCLEF LPSLPMAAF CGNMFVEPGE
451 QCDGFLDDC VDPCCDSLTC QLRPGAQCAS DGPCCQNCQL RPSGWQCRPT
501 RGDCLPEFC PGDSSQCPPD VSLGDGEPCA GGQAVCMHGR CASYAQQCQS
551 LWGPGAQPA PLCLQTANTR GNAFGSCGRN PSGSYVSCTP RDAICGQLQC
601 QTGRTPQLLG SIRDLLWETI DVNGTELNCS WHLDLGSDV AQPLLTLPQT
651 ACGPGLVCID HRCQRVDLLG AQECRSKCHG HGVCDSNRHC YCEEGWAPPD
701 CTTQLKATSS LTTGLLLSLL VLLVLVMLGA SYWYRRLHQ RLCQLKGPTC
751 QYRAAQSGPS ERPGPQORAL LARGTKASAL SFPAPPSRPL PPDVSKRLQ
801 SQGPAKPPPP RKPLPADPQG RCPGDLPGP GAGIPPLVVP SRPAPPPTV
851 SSLL

(SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION

N-glycosylation site

Number of matches: 5

- | | | |
|---|---------|------|
| 1 | 254-257 | NRTL |
| 2 | 406-409 | NFSN |
| 3 | 409-412 | NCSR |
| 4 | 623-626 | NGTE |
| 5 | 628-631 | NCSW |

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 11

- | | | |
|---|---------|-----|
| 1 | 53-55 | SEK |
| 2 | 73-75 | SLK |
| 3 | 199-201 | SWR |
| 4 | 283-285 | TQR |
| 5 | 411-413 | SRR |
| 6 | 589-591 | TPR |
| 7 | 602-604 | TGR |
| 8 | 611-613 | SIR |
| 9 | 686-688 | SNR |

FIGURE 2A

10 760-762 SER
11 796-798 SKR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 8

1 81-84 SLPE
2 199-202 SWRE
3 208-211 TPPE
4 283-286 TQRD
5 500-503 TRGD
6 522-525 SLGD
7 589-592 TPRD
8 611-614 SIRD

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 18

1 10-15 GLLGAG
2 145-150 GSWVSI
3 323-328 GTSFSG
4 358-363 GVASSI
5 404-409 GLNFSN
6 422-427 GMGSCL
7 475-480 GAQCAS
8 532-537 GQAVCM
9 555-560 GAQPAA
10 571-576 GNAFGS
11 583-588 GSYVSC
12 596-601 GQLQCQ
13 624-629 GTELNC
14 637-642 GSDVAQ
15 670-675 GAQECR
16 682-687 GVCDSN
17 714-719 GLLLSL
18 774-779 GTKASA

[5] PDOC00016 PS00016 RGD
Cell attachment sequence

501-503 RGD

FIGURE 2B

[6] PDOC00021 PS01186 EGF_2
 EGF-like domain signature 2

690-701 CYCEEGWAPPDC

[7] PDOC00129 PS00142 ZINC_PROTEASE
 Neutral zinc metalloproteinases, zinc-binding region signature

362-371 SIAHELGHSL

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	25	45	1.602	Certain
2	144	164	0.925	Putative
3	317	337	1.237	Certain
4	430	450	0.768	Putative
5	547	567	0.601	Putative
6	640	660	1.243	Certain
7	711	731	2.394	Certain

BLAST Alignment to Top Hit:
 Alignment to top blast hit:

>CRA|335001098640323 /altid=gi|7451525 /def=pir||G02390 disintegrin
 and metalloproteinase MDC15 (EC 3.4.24.-) - human
 /org=human /taxon=9606 /dataset=nraa /length=814
 Length = 814

Score = 1714 bits (4390), Expect = 0.0
 Identities = 812/855 (94%), Positives = 812/855 (94%)
 Frame = +1

Query: 76 MRLALLWALGLLGAGSPLPSWPLPNIALLSIPSVLSWGLGPAGGTEEQQAESKAPREP 255
 MRLALLWALGLLGAGSPLPSWPLPNI GGTEEQQAESKAPREP
 Sbjct: 1 MRLALLWALGLLGAGSPLPSWPLPNI-----GGTEEQQAESKAPREP 43

Query: 256 LEPQVLQDDLPISLKKVLQTSLEPLRIKLELDGDSHILELLQNRELVPGRPTLWYQPD 435
 LEPQVLQDDLPISLKKVLQTSLEPLRIKLELDGDSHILELLQNRELVPGRPTLWYQPD
 Sbjct: 44 LEPQVLQDDLPISLKKVLQTSLEPLRIKLELDGDSHILELLQNRELVPGRPTLWYQPD 103

Query: 436 GTRWVSEGHTELECCYQGRVRGYAGSWWSICTCSGLRGLVLTTPERSYTLEQPGDLQGP 615
 GTRWVSEGHTELECCYQGRVRGYAGSWWSICTCSGLRGLVLTTPERSYTLEQPGDLQGP
 Sbjct: 104 GTRWVSEGHTELECCYQGRVRGYAGSWWSICTCSGLRGLVLTTPERSYTLEQPGDLQGP 163

FIGURE 2C

Query: 616 PIISRIQDLHLPGHTCALSWRESVHTQTPPEHPLGQRHIRRRRDWVTETKTVELVIVADH 795
 PIISRIQDLHLPGHTCALSWRESVHTQTPPEHPLGQRHIRRRRDWVTETKTVELVIVADH
 Sbjct: 164 PIISRIQDLHLPGHTCALSWRESVHTQTPPEHPLGQRHIRRRRDWVTETKTVELVIVADH 223

Query: 796 SEAQKYRDFQHLLNRTLEVALLLDTFFRPLNVRVALVGLEAWTQRDLVEISPNPAVTLEN 975
 SEAQKYRDFQHLLNRTLEVALLLDTFFRPLNVRVALVGLEAWTQRDLVEISPNPAVTLEN
 Sbjct: 224 SEAQKYRDFQHLLNRTLEVALLLDTFFRPLNVRVALVGLEAWTQRDLVEISPNPAVTLEN 283

Query: 976 FLHWRRRAHLLPRLPHDSAQLVTGTSFSGPTVGMAIQNSICSPDFSGGVNMDHSTSILGVA 1155
 FLHWRRRAHLLPRLPHDSAQLVTGTSFSGPTVGMAIQNSICSPDFSGGVNMDHSTSILGVA
 Sbjct: 284 FLHWRRRAHLLPRLPHDSAQLVTGTSFSGPTVGMAIQNSICSPDFSGGVNMDHSTSILGVA 343

Query: 1156 SSIAHELGHSLGLDHDLPGNPCPCPGPAPAKTCIMEASTDFLPGLNFSNCSRRRALEKALL 1335
 SSIAHELGHSLGLDHDLPGNPCPCPGPAPAKTCIMEASTDFLPGLNFSNCSRRRALEKALL
 Sbjct: 344 SSIAHELGHSLGLDHDLPGNPCPCPGPAPAKTCIMEASTDFLPGLNFSNCSRRRALEKALL 403

Query: 1336 DGMGSCLFERLPSLPPMAAFCGNMFVEPGEQDCGFLDDCVDPCDSLTCQLRPGAQCAS 1515
 DGMGSCLFERLPSLPPMAAFCGNMFVEPGEQDCGFLDDCVDPCDSLTCQLRPGAQCAS
 Sbjct: 404 DGMGSCLFERLPSLPPMAAFCGNMFVEPGEQDCGFLDDCVDPCDSLTCQLRPGAQCAS 463

Query: 1516 DGPCCQNCQLRPSGWQCRPTRGDCDLPEFCPGDSSQCPPDVSLGDGEPCAGGQAVCMHGR 1695
 DGPCCQNCQLRPSGWQCRPTRGDCDLPEFCPGDSSQCPPDVSLGDGEPCAGGQAVCMHGR
 Sbjct: 464 DGPCCQNCQLRPSGWQCRPTRGDCDLPEFCPGDSSQCPPDVSLGDGEPCAGGQAVCMHGR 523

Query: 1696 CASYAQQCQSLWGPQAQPAAPLCLQTANTRGNAFGSCGRNPSGSYVSCTPRDAICGQLQC 1875
 CASYAQQCQSLWGPQAQPAAPLCLQTANTRGNAFGSCGRNPSGSYVSCTPRDAICGQLQC
 Sbjct: 524 CASYAQQCQSLWGPQAQPAAPLCLQTANTRGNAFGSCGRNPSGSYVSCTPRDAICGQLQC 583

Query: 1876 QTGRQTPLLGSIRDLLWETIDVNGTELNCSWHLDLGSQVAQPLLTLPGTACGPGLVCID 2055
 QTGRQTPLLGSIRDLLWETIDVNGTELNCSWHLDLGSQVAQPLLTLPGTACGPGLVCID
 Sbjct: 584 QTGRQTPLLGSIRDLLWETIDVNGTELNCSWHLDLGSQVAQPLLTLPGTACGPGLVCID 643

Query: 2056 HRCQRVDLLGAQECSRKCHGHGVCDSNRHCYCEEGWAPPDCTTQLKATSSLTTGLLLSLL 2235
 HRCQRVDLLGAQECSRKCHGHGVCDSNRHCYCEEGWAPPDCTTQLKATSSLTTGLLLSLL
 Sbjct: 644 HRCQRVDLLGAQECSRKCHGHGVCDSNRHCYCEEGWAPPDCTTQLKATSSLTTGLLLSLL 703

Query: 2236 VLLVLVMLGASYWYRRLHQRLCQLKGPTCQYRAAQSGPSEPPGPPQALLARGTKASAL 2415
 VLLVLVMLGASYWYRRLHQRLCQLKGPTCQYRAAQSGPSEPPGPPQALLARGTK
 Sbjct: 704 VLLVLVMLGASYWYRRLXQRLCQLKGPTCQYRAAQSGPSEPPGPPQALLARGTK---- 759

Query: 2416 SFPAPPSRPLPPDPVSKRLQSQGPAKPPPPRKPLPADPQGRCPSGDLPGPAGIPPLVWP 2595
 SQGPAKPPPPRKPLPADPQGRCPSGDLPGPAGIPPLVWP
 Sbjct: 760 -----SQGPAKPPPPRKPLPADPQGRCPSGDLPGPAGIPPLVWP 799

FIGURE 2D

Query: 2596 SRPAPPPPTVSSLYL 2640

SRPAPPPPTVSSLYL

Sbjct: 800 SRPAPPPPTVSSLYL 814 (SEQ ID NO:4)

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF01421	Reprolysin (M12B) family zinc metalloproteas	259.3	5.3e-74	1
PF01562	Reprolysin family propeptide	128.4	2.1e-35	1
PF00200	Disintegrin	70.0	3.4e-22	1
CE00385	E00385 platelet_aggregation_activation_inhib	26.5	5.4e-06	1
PF00035	Double-stranded RNA binding motif	7.2	1.2	1
CE00423	E00423 stromelysin_1	4.5	0.99	1
PF01400	Astacin (Peptidase family M12A)	2.6	7.8	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF01562	1/1	100	217 ..	1	119 []	128.4	2.1e-35
PF01400	1/1	363	373 ..	91	101 ..	2.6	7.8
CE00423	1/1	364	375 ..	222	233 ..	4.5	0.99
PF01421	1/1	230	428 ..	1	200 [.	259.3	5.3e-74
CE00385	1/1	447	518 ..	1	67 [.	26.5	5.4e-06
PF00200	1/1	447	523 ..	1	76 []	70.0	3.4e-22
PF00035	1/1	734	766 ..	1	37 [.	7.2	1.2

FIGURE 2E

1 TTGGGTGACC CTGGGCAGTG ATCACATCTC CAAGCATCAG TTTTCTCACC
51 TGAAAAAAG GAGATGATAA TAACACTATC TGCCTTACAT GACAATTGAA
101 TTGAATTTTT TTTTTTTTTT TGAGACTAAG TCTCACTCTG TCGCCCAGGC
151 TGGAGTGCAG TGGCGTGATC TTGGCTCACT GCAACCTCCA CCTCCCCAGT
201 TCAAGCGATT CTCGTGCCTC AGCTTCCCGA GTAGCTGGGA TTACAGGCAC
251 AACTACCAC GCCCGGCTAA TTTAGAATTG AAATAATTTA TGTACAGTAT
301 CTTAGTACAG GACCTGACAT TATAAACAAAT GAGTGGCAGC CATTCTTATT
351 TAATCAGTCC TAACAAAGTT CATAAAAGTG AGACTGTGTT TGCTTAGCTT
401 TTTCCCTAGG GCCTGGATAC CCCCAGCCCC CATGACACAC AATAGGGGCC
451 AAATGAATGT GTTGTGAAAA AATGAAAAAC AAAAAACAAA AAAGAACATG
501 CTGGGATTCC TTGACAGGGT CGTGAAGCAA ACTGAATGTG AATGCACAGA
551 TGGAAATGTG CCAGACAGTC ATTCCAAGCA GAATGTGCAA AGACTCAGTC
601 CACAGGGAAT GCGAAGTGCC AGGGCTAGTC TCAGGAGAAA CTGGCTCAGA
651 AGAGACAGCT CTCAGGGAGG GCTAAAGTAG GAAAGAGGCT AGAAAGGGAC
701 CAGGTGAGGG AAGGCTCTGA AGGCCAAGCC CAAGAGTTCT GCCTGTCTGG
751 CAGGCAGCAG GGCCTCTGGA GTTCTTGGG CAAAGAGTGG CTGCTTCCTG
801 GGTAAGGTGG CCTGTGGAAT ATCCCTGACA ACTGTGTAGA GACATGTCGT
851 GAGGGATGGC AGGGAGCATA GTGAACTAGG TTTGTGGTTT GGAATCAGGG
901 CCCCTGGGGT CCAGCCAAGT TGGATTGTTT ACTATCTGTG TGACTTTGAG
951 AGTCACTTCA CCTTTCTCAA CTGTAAAGTG GGGATAGCAA CAGTGATAGT
1001 CGATCTGGCC TGCTCACTTC TCAGCCTCAC TGTGAGAACC AAATAAGATG
1051 ATTTACAGGA AAGTGCAAAT GAGAGTTGTG GCTGATATCC GCTTGGAGAG
1101 AGCCTGGAGG GTGCATCCTC CCATTCTCCA TCACAGAGTT GGGGAGGGAG
1151 GCACCCTCGC CCTCCAGGGG TTTCTTTTGT CCAACCCAGC CTCCTCCAAC
1201 ACGCGGGAAT TGTGAGCCTT GCGGACTTCA GACAGGAAAC GCTGTCCAGT
1251 TCCCTTTCTT TCCCGCCTCG CTCCCGGGCT GCGGCTAACG CCCACCTCCC
1301 AACAGCGCCA CCCGCTGGCG GATATCCTGC ACCGCGGCTG CCCGCTCCTG
1351 CGCCGCTGGC TGTGCCGCGC CTGCGTGGTG TGCCAGGCAC CCGAGACGCC
1401 CGAGTCTTAC GTGTGCCGGA CGCTGGACTG CGAGGCCGTG TACTGCTGGT
1451 CGTGCTGGGA CGACATGCGG CAGCGGTGCC CCGTCTGCAC GCCCGCGGAA
1501 GAGTCTCTT CCTCCGCTT TAGTGACAGC AACGACGACA CTGCCTACGC
1551 GGGGTGAAGA GCGTCCTGC TCGCTCTTCC GCACCGTCTT TCCCGTTAA
1601 TAAATGCCC TGTACGCTT ACGTGGGTCG GGGACTGGGG TGAGCCGCGC
1651 ACTGCCTCGC CTGCAGTCGG GAAAGCCTGC CCGCCCGACC TCTCCGAGCC
1701 AGGCCGCGCA CAGGAGGCAG GGAGGCCGCG AAGCTACTAG GGAGGGGTCC
1751 GGACCTGGCG CCGGGTGAAG GCGCGCCGCC CAAGCCGGTC GGACCGGGCA
1801 CCGGCTCCCA CTCCGCACAG TTGCGGGGAA GCGGTAGCGC TGAGCAGCGC
1851 GGGCGTAGTG GCGGTGTCC CCGTCCCGA GGCACCCGGC GCGCAGCGGG
1901 GCGGGCTTTG CCGGGGCGG AGCTTGGCTT GGGGCCGGGT GGGAGGGGGC
1951 GGGCCGGGGC GGGGCTGTT GCGCGCGCG CGCTGCTGGG TTCTCCGAGG
2001 CGACCTGGCC GCCGGCCGCT CCTCCGCGCG CTGTTCCGCA CTTGCTGCCC
2051 TCGCCCGGCC CGGAGCGCCG CTGCCATGCG GCTGGCGCTG CTCTGGGCCC
2101 TGGGGCTCCT GGGCGCGGGC AGCCCTCTGC CTTCTGGCC GCTCCCAAAT
2151 ATAGGTGAGT CCTCCGCTG GAGTGGGTCG GGGGGCGGAC TGGGAGGGAG
2201 GTGCAGGAAA GTCGGAAGGC ATTAGGGTAA TGGGGCCGGA CGGAGACCCT
2251 GGGAGAGCCC AGCCAGAGCG CGGCCCGCCC TGGTCCGCTG TCCTGGGCCT

FIGURE 3A

2301 AGGGCCCCGT GACTTGGCGA TGGGGTGAAA AGAGAAGGAG GGGGGATGCC
2351 GGCGCCCCCT GCCTCCTGCC TGGTCATCCT CTGCGCGGTC CCTGCGGACA
2401 CTTTCAGGCT CAGGTACCAG GTACCGAGGG GCCTGTCCAG CGCCACTTCA
2451 AGATCGTGAT GAGAGGGTCG CTGCTCCCCA GGA CTG GGCAT CTTGCTGCT
2501 CTGGGGCCTA GCTAACCGTT CCACCCGGTG CCAGGGCGCT GAGCGGGCAT
2551 GGCTTGTAGG GTTTAGTGAA GAGGATTCTC TCTAGCCTCT ATTCCAGGCC
2601 TGGGGCCGCC AGGCACTCCT CACCCTGGTG CTGTTGCCAC CAGTGCCTGG
2651 CCGAGCGGGA GGGGCCCAG ATGAGCCAGG AGAAGGGAGA ATTGGCCAGG
2701 AAAGAGGCTG GGACACCAAC TCCTCCTTGG AACTTTCACT TCCCGCTGCT
2751 GTCTTGGGCC GGGACCGAGA GGGCAGGCGC GGGTGGAGTG TCCGAGGAG
2801 AGAGGGCCAT TGTGTGTTGG GGGGGTGGGG GGTGCTCGAG GAGGAAGCAG
2851 AGGCTGTAGG CAGCGGGTGT GCCTGACTGG GCATGAGGGT GTTTAGGGAG
2901 GTGGGGGTGT TTGCACTGCT CACCCAGAAA TGGGCGTTCC TGGCATCTCC
2951 GATGTGAGCG AAGGGGAGGG TGAGCGGGCA CCCGGCCACA AGGCTTAGCT
3001 CAGTCTCGAG AGGGGGCGTT CCTGAAGTGG GGGGAGAGTG ATTGGGAGGG
3051 AGTGGGAACC GCGGAGGGTC CTGTGAGAAC CTGGGATTGG CCGGAAGGGG
3101 ACAAGGAGGG CCACAGGCTG CGCAAGCCGA AAGTCTTTCT TGGGGA CTTG
3151 TGAATGGGT GGTGGGTGGA AAGCCATAAA TTAGAGAGAC ACCCTCTCCT
3201 TCCAGTATTC TTCTTTAAGT CTCAGCATGC AATGTGGAAG CCCCTCAGGT
3251 ACCTAAGGGT CTTGATGGGC TGGGAGCTGG TGGATCTGAG GGCACCTGTC
3301 ACCCCCAGCC CTGCTGTGCA TTCCCTCAGT ACTGTCTTGG GGTGTCCTGG
3351 GACCTGCAGG TGGCACTGAG GAGCAGCAGG CAGAGTCAGA GAAGGCCCCG
3401 AGGGAGCCCT TGGAGCCCCA GGTCTTCAG GACGATCTCC CAATTAGCCT
3451 CAAAAAGGTG CTTCAAGGTGA GCTCTCACTC CCCTCTAATA AATAAACGAA
3501 TCCACACACG CCCCAGTATA GCCAGGTGTC TCAAAGCCAA AGCTTGGCTG
3551 AGGAGCTGGT GGGTAGAGCT CACTGTAGTG GGTCTATCCC AGGCCAGCT
3601 GCCTCTCCA CCACACCCA GCACCTGGCT TCACTTATCT CCCTCTCCCT
3651 CTGCACACAC GTGTATCTGT CTGCCTCAGC CCCACCCAAC CCATCCATCT
3701 CCACTGGGGA AATTGTGAAG CCAAAC TTGC TTTCTTCATC TCATGTTGTC
3751 GGTTTTCTCA GTGGGGGGAT TTGGAAAGAG TCAGGACCTT ACCAAACCCC
3801 CCCCCCCCAC CCCATTCTAA AGCTGAGTCA GAGGAAGGGC TGGGGCTTGT
3851 GCTGGGTCCT ACACGGTGCT TCCTCTCTGG GCAGGAAGCC GAGAAGGGGT
3901 GGCTCAGATA CCTTCCTTGA CCTCCGCACA CAACCCCCCA GAACAATGCT
3951 CCAGGCCAGG CAGGGTTTCC TGGCCCCTCC CCTGGGATCC CCCACCAGT
4001 GATCTAATTG CTGGTGCTCT TCTGTGGGCC TGAGGTTTTT TGGTTAGAGA
4051 GGCTGGGAGT TGTGGACAGG TCTAGGGAGG TGACCTGCCC TCTGGTGCCC
4101 ACAGACCAGT CTGCCTGAGC CCCTGAGGAT CAAGTTGGAG CTGGACGGTG
4151 ACAGTCATAT CCTGGAGCTG CTACAGAATA GGTAATAGTG ATGGTGCCAA
4201 TAACAGTGAC CACATGGCCA ACAACTTGTA TAGCATTTAT TATGTGCCAG
4251 GACTAAGTG CTTGTGCTCA TTTAATCCTC ATAACAGCCC TATAAGGGAT
4301 ATACTATCAT GTATTATTGT CCTCACTTTA TACATGAGGA AGTCAAGGCA
4351 CAGAGAGATT AAATAACTTG CCCCAGGTCA CACAGCTAGT ATGTGGTGAA
4401 AACCAGATTG GAATTCAAAT AAATAACAG AGTCAGTGGC CCAACCAGTA
4451 TACTTTGCTG CCCCAGGGTC AGGAGTGGAA AAGTTGGCTG CGGGGGTTGC
4501 CTGGTCCCCA GCCCCACAAC CACCTTCAAG CCTCTGCTTG TCAATGCACC
4551 GACCCTGGGA AGTGGCTTTA GCACTGCCTT CTTTTCTTC ACTTCACAGG

FIGURE 3B

4601 GGAGTTGGTC CCATGTCCGC CCCGACCCTT GGGGTCCGGC TNTCCCCTCT
4651 CCCCCCTTCG GCGCCGCCCC TTCCCTTTTC TTTCTTCCCC TCCGCTTTCG
4701 TCCTTTTGCC TCCCCCGTGC CGTTGCGCGT TCCTTCTTCC CCGTTCCCTC
4751 TCCCCTCTTT TGTTCCCTCC CGTTCTTTTC TCCCCCGCGT TCTTTCCTCC
4801 TCCTTTTCGG TCCGCCCTCG CCTTCTCTCC TTCCCCTTCT GCCCTTCGCC
4851 NTTTCTCCCT CTCGTTCTTC CTCGGTGTGC CGTCGTCCCG GCTCGGCCTT
4901 TCCCCGCTTC CTCCCGCTCG CCGTTTTTTT CCCCCCGCTG TCTTCCCGTG
4951 TTCCCCTTCG CTTCTCCTCT TCCCTTTCGT TCGGTCGTTT TCTCGTTCCA
5001 TTCCCGCCTC CCCGTTTCCG TTCCACTCCT TCTTCTCCT TTCCCGCTCC
5051 CCGTTTCTCC CGACCCCAAC AACAAATAAA NNNNNNNNNN NNNNNNNNNN
5101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNTCAGG
5501 AGGCCGAGTG GAAGAATCGC TTGAGCCAG GTAGGCAGAG GTTTCAGTGG
5551 GCCGAGATCG AGCCACTACA CACCAGCCTG GGTGAAAGAG TGAGACCTCG
5601 TCTCAAAAAA TAAAATAAAA ATAAAATAAA ATAAAATCTA GCTGAGACAG
5651 ATTAGGTGGT TTGCCCAGAG CCTACAAC TATAAATGGC CTATCCATTT
5701 ATTAGTTGTA TTTGGCTCTT CATCTGTCTT ATGATCCCAT TTGCAGAGAG
5751 CTCTCACTTG GTTATAGATA ATACATAGTT ACCAATGATG AAGCAATATA
5801 AACCCAATTT CCTAATTTGT AAAATGAAGA TAATAAACT ACTTGCTGCA
5851 TAGAGTTGCT GGAAGATTA AATAAGTCCA TATAGATGTA AAGTGCTTAA
5901 AACTATGCCA GACCTATGGT AAGTGACAAG AGTTGTTATT GGGATTTTTA
5951 AAATTATTAT TATTATTATT ATTATTATTT GAGACAGAGT CTCGCTCTGT
6001 CTCCCAGGCT GGAGTGCAGT GCGGTGATCT CGGCTCACTG CAAGCTCCGC
6051 CTCCCAGGTT CACGCCATTC TCTTGCCTCA GCCTCCCGAG TAGCTGGGAC
6101 TACAGGCGCC CGCCACTACA CCCGGCTAAT GTTTTGTATT TTTTAGTACA
6151 GACAGGGTTT CACCGTGTTA TCCAGGATGG TCTCGATCTC CTGACCTCAT
6201 GATCCACCCG CTTGTCTCTC CCAAAGTGCT GAGATTACAG GCGTGAGCCA
6251 CCGCACCCAG CTAAATTAAT GTTTTTTAAA AATTTGAAAA AAACCACTGA
6301 GTTTGGAGCC AGAAAAGCAG GGGTCTACTC CAACCTTCAT TATCTACTTC
6351 CTGGTCTCTC TTGGCAAGTT CCTGGGCCCT CTGGCCTTCA GTGGCTCATC
6401 TGTAATATGG GCTCTTCACC CTCCTATTTG ACCACAGAG TAGGAGTGGC
6451 TGCCTCTTGG TCAGCCCGGC ACAGCTGCTG GCTGCGAGCG GCAGGTTTGC
6501 CTGATAATTC TTCTTGTCCA TAGTAGAGGC GGGATGTGGT AACAGAGACC
6551 AAGACTGTGG AGTTGGTGAT TGTGGCTGAT CACTCGGAGG TGAGCCTGCT
6601 GGCCCCCTGCA CATCTCCTC CCCCTGCACT GCCCTGCCGC CTTTCATGTC
6651 ACCTCTCTTG GCCTACAGGC CCAGAAATAC CGGGACTTCC AGCACCTGCT
6701 AAACCGCACA CTGGAAGTGG CCCTCTTGCT GGACACAGTG AGTGCTGGAC
6751 AGGGCAACCC CCACCCAGG CCCCTGACCA TGGCAACCC TCTTCTGAGC
6801 CCCAGCTGTC TTTCACTTCT TCCGGCCCCT GAATGTACGA GTGGCACTAG
6851 TGGCCTGGA GGCCTGGACC CAGCGTGACC TGGTGGAGAT CAGCCCAAAC

FIGURE 3C

6901 CCAGCTGTCA CCCTCGAAAA CTTCCTCCAC TGGCGCAGGG CACATTTGCT
6951 GCCTCGATTG CCCCATGACA GTGCCCAGCT GGTGACGTAA GGGCCCCAGA
7001 CTCAGCCAGA GAGGCCAGTC CTGTCTGGC CAAATTCACA CCCCTTCAGC
7051 ACCCTACCTC AGCCCCTGAA GCTCTGACCA CCGTGGCTTC TGGCCCTGAA
7101 CTTTAGCCTC TCTGTCCAC AGTGGTACTT CATTCTCTGG GCCTACGGTG
7151 GGCATGGCCA TTCAGAACTC CATCTGTTCT CTTGACTTCT CAGGAGGTGT
7201 GAACATGGTG AGTTATTTCC AGGTCTCCTC CTCATTCCA ATTCAGTTCC
7251 TCCCAAGTGT GGTGGCATT TATGCACTGAA ACCCCCCTAT AAAGTTGCCC
7301 AACCCCAAAG CTACAGGTAT AGAGGGTGA GGTACGTGAT GTGGCCTTTG
7351 CTATCAGGGA GCCCTCGCTT ATGGCCAGCT AGTCACAGTG TACACAGTCA
7401 TCCCCTGTGC AGTCTTCCA TTTCTTAGAG GAGGGTAGGA GGCAGCTAAG
7451 GCCCAAAGAA CAGAGGTGAT CTCCCTCCAG TGAGGGAGGG GGACAGAGCT
7501 GAGCTAGAAC CCAAGTTTCT GCCATCCAGG CCTGGGTTCT CCTACTTTAG
7551 AAGCAATTCA GGAGGGAAGC AGTGCCTGCT GAGTGCCAC GAGGTCAGAC
7601 GTGGAGGGAA CAGGAGCAGA GAGGGTGGT TGGGCATTGT GGTGGAGGCA
7651 GGCTGGGACT GGACCTACAG TACCCCTCCC CAATGACAGG ACCACTCCAC
7701 CAGCATCCTG GGAGTCGCCT CCTCCATAGC CCATGAGTTG GGCCACAGCC
7751 TGGGCCTGGA CCATGATTTG CCTGGGAATA GCTGCCCTG TCCAGGTCCA
7801 GCCCCAGCCA AGACCTGCAT CATGGAGGCC TCCACAGAGT AAGTAGCTGC
7851 AGGATGGAGA GAGGGTGTGG GGCAGGGGGC AGGGANNNNN NNNNNNNNNN
7901 NNNNNNNNNN TGTTAGAGTT ACCTTCCTTG CCACCTCCC CAGCTTCCTA
7951 CCAGGCCTGA ACTTCAGCAA CTGCAGCCGA CGGGCCCTGG AGAAAGCCCT
8001 CCTGGATGGA ATGGGCAGCT GCCTCTTCGA ACGGCTGCCT AGCCTACCCC
8051 CTATGGCTGC TTTCTGCGGA AATATGTTTG TGAGAGCCGG CGAGCAGTGT
8101 GACTGTGGCT TCCTGGATGT GAGCCCTTT CCCAAAGCCT CGCCCCACTC
8151 ACTTCTGTAC CCTCACCTG GCTCATTAGC CCTATCCCAG CCTCCTGAGC
8201 TCTTGGGTTT TGAAGGGACT TTCCACCCCT CTCCTACTTG CCCTGTCTGT
8251 GGGGACAGCA CATGGGTTGT TGGGCTCTAG CCCTCGCTTG CTGTGTAGCT
8301 TCTGGTCTTG GCCTGTGGGA GGAGGAGAGA TTGGAGGGAG GCTCACAGGC
8351 CCCACCTGCT CTGATGCCCG GCCCCGTGC TCCTGCCAC AGGACTGCGT
8401 CGATCCCTGC TGTGATTCTT TGACCTGCCA GCTGAGGCCA GGTGCACAGT
8451 GTGCATCTGA CGGACCCTGT TGTCAAAATT GCCAGGTGGG TAGAGACTAG
8501 ACTGGCCACC CGGAGCTCAC CTGCCGGGGC CAAGGTGGAA AGGGTCATTC
8551 TGACCCCCGG CTGGATTTGC TCAGTGCCCA CACTGATGCT CATCCACCCT
8601 CCACAGCTGC GCCCGTCTGG CTGGCAGTGT CGTCCTACCA GAGGGGATTG
8651 TGACTTGCTT GAATTCTGCC CAGGAGACAG CTCCCAGTGT CCCCTGATG
8701 TCAGCCTAGG GGATGGCGAG CCCTGCGCTG GCGGGCAAGC TGTGTGCATG
8751 CACGGGCGTT GTGCCTCCTA TGCCAGCAG TGCCAGTCAC TTTGGGGACC
8801 TGGAGCCCAG CCCGCTGCGC CACTTTGCCT CCAGACCGCT AATACTCGGG
8851 GAAATGCTTT TGGGAGCTGT GGGCGCAACC CCAGTGGCAG TTATGTGTCC
8901 TGCACCCCTA GGTAAAGTGA GAAACCTGGC TCCTCCTTTG GGTTTCTGAG
8951 AGCCTTGCC CTGCTCCTAC TAACTCTGTG TGCCCTTCCC CCTCNNNNNN
9001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNTTACGG
9051 CATTTGTAGT TACTCACACT TTTGCCTTCA NACAGCTAAT ACTCGGGGAA
9101 ATGCTTTTGG GAGCTGTGGG CGCAACCCCA GTGGCAGTTA TGTGTCCTGC
9151 ACCCCTAGGT AAGTGAGGAA ACCTGGCTCC TCCTTTGGGT TTCTGAGAGC

FIGURE 3D

9201 CTTGGCCCTG CTCCTACTAA CTCTGTGTGC CCTTCCCCCT CCCCACAGAG
9251 ATGCCATTTG TGGGCAGCTC CAGTGCCAGA CAGGTAGGAC CCAGCCTCTG
9301 CTGGGCTCCA TCCGGGATCT ACTCTGGGAG ACAATAGATG TGAATGGGAC
9351 TGAGCTGAAC TGCAGCTGGG TGCACCTGGA CCTGGGCAGT GATGTGGCCC
9401 AGCCCCCTCT GACTCTGCCT GGCACAGCCT GTGGCCCTGG CCTGGTGAGC
9451 AGCCTGGGTG GGCAAGACCA GGTGTGAGAA GGGACATTTG GACCACAATG
9501 AACAGAGCCC AGACTTCACC ATTACCAAT GTCAAAGGCA GGGACTCCAA
9551 GGAAGTCAG TTTCTTACTT CAGATGGAGC AAAGTCCTAT CAACTCACTA
9601 TGCCTTGTT TCCCCATCTG TAAACGCAGG GTATGGCCTC AACCTTATTG
9651 GCCTCCAGT CCCATTAAAG CTTTGTGGGA ATCTGATCCA GGCTCTTCTC
9701 TCCCTGGGTC AGGTGTGTAT AGACCATCGA TGCCAGCGTG TGGATCTCCT
9751 GGGGGCACAG GAATGTCGAA GCAAATGCCA TGGACATGGG GTGAGCTGGG
9801 ATGGGGGAAG TGAAGGGGA GCAGAGAGCC TCTAGAGAGG AAAAGGATAC
9851 TGGGCTTTGG AAATAGACAT ATCTGGGTTT TAATCCTTGC TCTACTACTT
9901 CCCAGTTGTG TGACCTCGGG CAGGTTACTA ACTTTGCTGA GCTCAGTTTC
9951 CCCACCTATC AAATGGCTAT AATAATAGTA TCCCCATCCA GGGTACATGA
10001 GATGTGTATG CAAGCAAGTA GCACAGTGGG TAACTAATAG TGCTTTTAAA
10051 AAAAAAAAAA NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3E

FIGURE 3F

13801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14751 NNNNNNNNNN NNNNNNNNNN NNNNNNTTTT TGAAAGCTAC TAGTAGGTCA
14801 CCATTTTTTC TTGTCTTCCC GCAATCCAGA CCAGCGCCAC CGCCTCCGAC
14851 AGTGTCTCG CTCTACCTCT GACCTCTCCG GAGGTTCCGC TGCCTCCAAG
14901 CCGGACTTAG GGCTTCAAGA GGCGGGCGTG CCCTCTGGAG TCCCCTACCA
14951 TGA CTGAAGG CGCCAGAGAC TGGCGGTGTC TTAAGACTCC GGGCACC GCC
15001 ACGCGCTGTC AAGCAACACT CTGCGGACCT GCCGGCGTAG TTGCAGCGGG
15051 GGCTTGGGGA GGGGCTGGGG GTTGGACGGG ATTGAGGAAG GTCCGCACAG
15101 CCTGTCTCTG CTCAGTTGCA ATAAACGTGA CATCTTGGGA GCGTTCCCCA
15151 GAGTTTGTCT GCTTCTAGAA CCCGGGTCGC TCCTGCTGCG GTTCCAGGTT
15201 TGGCCGCCAG AAGACGCTGC CGCCTCAGAC GAGGGCGGGC TGTGTGGGGC
15251 GGGAGTACCA GAAAGGGTCG GCGTGTGTCC CCGGATGCT CGCAGCTTCC
15301 CTCTGCCAG ACTGGGGTGG CTTTCGGCGC AATCTGTCAA GCTGTTGGAC
15351 CTGCCGTCCC CACTCTGACC ATTGGCTGGG AAAAGTGGAT CTGGCTGATG
15401 CTCCCAGAGC CCAGGAGCCA GGGCGGAGCG GGGCGGCGGC TGCTCCACG
15451 ATCCCAAGGC CGCGCACCTG CCTCCTCCCC CTCCGCCGCC GCCACTTGAG
15501 GGATCGGGAA CAAAGGTGCT TTGTACAGGC CGCAACCACC TCATTACTTC
15551 GTCTTAGGGA CTGGGGCCGC GTGGGCCCCC AGCCC GAAC GAAGGTGTGG
15601 AGCGGCAAGG GACAGACGCC AATCTTAAAG TGAGCATCTA GCGCGCCACC
15651 TAAGGCTCTT TAGGGAAGGT GGTCCCAGAG CTGTGTTGTC CCTTCCGCTT
15701 GCACTGTCCC TAGATGTGCA AAGAAAACGG GGCAGTGCAT GAAGGTGGTT
15751 GGACAGGCTT CATGGATCCT CGCCCGCGCC TCACTTTCCC CTATCTGGGC
15801 AAAGGTTATG TACCCTTATT TAAAATCTTC CAAACTTCTA ATAAGGCAGT
15851 CTACCCTGCA CTAAAGCAGA CACGAAAGAG ATGACCTCCC TAAAAATACT
15901 GCTGTTGGAA TACGTCCTTC CTTCCCGCCC CCTCGCAGTG CGGTGCAGCC
15951 TCA GTGGAAG CTTTGGCGAA CCTGGCGCGC GCTGCGGTGC ACAGAGGGTT
16001 AACTGGAGTT GGCGCTGGGT GGAGAGGAGG AGACGCGCTC CCATTGGCGG
16051 AAAGTTATTC AGGGGCGGGG TCAGTGAATC TCCGTACCCC ACTCCCCTTT

FIGURE 3G

16101 CCGCAACTTC CCTCTTCACT TTGTACCTTT CTCTCCTCGA CTGTGAAGCG
16151 GGCCGGGACC TGCCAGGCCA GACCAAACCG GACCTCGGGG GCGATGCGGC
16201 TGCTGCCCCCT GCTGCGGACT GTCCTATGGG CCGCGTCCTC GGCTCCCCCTC
16251 TGC GCGGGGG CTCCAGCCTC CGCCACGTAG TCTACTGGAA CTCCAGTAAC
16301 CCCAGGTAGC CGGGCCGAAC CGGGCGAGCG CACAGCCAAG TCTGCGCGCT
16351 CCCGGGCTTT GCGCGCGCCC GCCACCCGCT CTTTGCGCGG CGCCGCCTGA
16401 GCCTGGCCGC GCGCCGGGGC TCCTTTGT TT GAGCCGGCGG GGGAGGGGGG
16451 AGGGGCGAGG GCGGAGGCGC GCCCTGGGTC TCCCCACAGC CCGCATGTGT
16501 TGGGGGGCAG GCAGAAGACC CCAGCCCCAA GGGTTGTCTA GGGGGTCTTG
16551 GAGCATGGAG CTGGGGGGGC CTTTGCCCGC ACTCCGGGCT CCGCCCCCTC
16601 CGCTGCTCTC CTGGCGATCC CCAGCCTCCC GCAGGCTGGA GCTGTGGCTG
16651 ACGAATTGA GAGCGAGGGA GGGGGCTTTA CTCTTATGAA AGAGCGTGGG
16701 TTA CTCTCT GCCCCGTGGG TCTCACCTCT GGCTCTCACT CTGTCTCCTG
16751 ATCTCATTTG CTATCTCTGC TTTCATCTCT GTCTTTATTG GTCCTTCTGT
16801 TTCTTTCCAG TGTCAGCCCT GCCCTTCTAG CCGAATCACC TCTGGGCAAG
16851 TCTCGTGACC TTCCTAACCT CATTTATCTC ACCTGTATAA TGGGCTAATA
16901 ATACCTAGTA CCCTGGGAAG TCTGGCAGGG TAAGTGAGGT CATGTATGTG
16951 AAAGAGGCTC AGGCTGTACA GATATAAACT ATTATTTCTT TCTCTCTCCT
17001 GAGCTGCCTG CCTTTGAACC TTAGTATATT TTA CTGTTTC CATCCCCCTC
17051 CCCAAGTCTC CCGCTCTC CTATTTCTTA TCTGTTTTTC TTTCTGATTT
17101 TCTACTTGAG ACAATCTGTG ACTATTCATT TCTTCACT

(SEQ ID NO: 3)

FEATURES:

Start: 2076

Exon: 2076-2154

Intron: 2155-3308

Exon: 3309-3466

Intron: 3467-4104

Exon: 4105-4181

Intron: 4182-4206

Exon: 4207-4250

Intron: 4251-4436

Exon: 4437-4607

Intron: 4608-5048

Exon: 5049-5052

Intron: 5053-6007

Exon: 6008-6145

Intron: 6146-6528

Exon: 6529-6589

Intron: 6590-6668

Exon: 6669-6737

Intron: 6738-6816

Exon: 6817-6986

Intron: 6987-7122

Exon: 7123-7207

FIGURE 3H

Intron: 7208-7689
Exon: 7690-7838
Intron: 7839-7943
Exon: 7944-8118
Intron: 8119-8392
Exon: 8393-8485
Intron: 8486-8606
Exon: 8607-8911
Intron: 8912-9248
Exon: 9249-9444
Intron: 9445-9712
Exon: 9713-9791
Stop 9792

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
2522	C	G	Intron			
4326	C	T	Intron			
5954	T	-	Intron			
6783	G	A	Intron			
7514	A	C	Intron			
15505	C	T	Beyond ORF(3')			
16123	A	G	Beyond ORF(3')			

Context:

DNA
Position

2522 TTAGGGTAATGGGGCCGGACGGAGACCCTGGGAGAGCCCAGCCAGAGCGCGGCCCCGCCCT
GGTCCGCTGTCTGGGCCCTAGGGCCCGGTGACTTGGCGATGGGGTGAAAAGAGAAGGAGG
GGGGATGCCGGCGCCCCCTGCCTCCTGCCTGGTCATCCTCTGCGCGGTCCCTGCGGACAC
TTTCAGGCTCAGGTACCGAGGTACCGAGGGGCCTGTCCAGCGCCACTTCAAGATCGTGATG
AGAGGGTCGCTGCTCCCCAGGACTGGCATCTTCGCTGCTCTGGGGCCTAGCTAACCGTTC
[C, G]
ACCCGGTGCCAGGGCGCTGAGCGGGCATGGCTTGTAGGGTTTAGTGAAGAGGATTCTCTC
TAGCCTCTATTCCAGGCCTGGGGCCCGCAGGCACTCCTCACCCCTGGTGCTGTTGCCACCA
GTGCCTGGCCGAGCGGGAGGGGCCCCGAGATGAGCCAGGAGAAGGGAGAATTGGCCAGGAA
AGAGGCTGGGACACCAACTCCTCCTTGGAACTTTCACTTCCCGCTGCTGTCTTGGGCCGG
GACCGAGAGGGCAGGCGCGGGTGGAGTGTCCGAGGAGAGAGGGCCATTGTGTGTTGGGG

FIGURE 3I

4326 GGGCCTGAGGTTTTCTGGTTAGAGAGGCTGGGAGTTGTGGACAGGTCTAGGGAGGTGACC
TGCCCTCTGGTGGCCACAGACCAGTCTGCCTGAGCCCCTGAGGATCAAGTTGGAGCTGGA
CGGTGACAGTCATATCCTGGAGCTGCTACAGAATAGGTAATAGTGATGGTGGCAATAACA
GTGACCACATGGCCAACAACCTTGATAGCATTTATTATGTGCCAGGTAAGTGCTTGT
GCTCATTTAATCCTCATAACAGCCCTATAAGGGATATACTATCATGTATTATTGTCCTCA
[C,T]
TTTATACATGAGGAAGTCAAGGCACAGAGAGATTAATAAAGTTGCCCCAGGTCACACAGC
TAGTATGTGGTGAACACAGATTGGAATTCAAATAAACTAACAGAGTCAGTGGCCCAACC
AGTATACTTTGCTGCCCCGGGGTCAGGAGTGAAAAAGTTGGCTGCGGGGGTTGCTGGTC
CCCAGCCCCACAACACCTTCAAGCCTCTGCTTGTCAATGCACCGACCCCTGGGAAGTGGC
TTTAGCACTGCCTTCTTTTCTTCACTTACAGGGGAGTTGGTCCCATGTCCGCCCCGAC

5954 AGGTGGTTTTGCCCCGAGGCCCTACAACATAATAAATGGCCTATCCATTTATTAGTTGTATTT
GGCTCTTCATCTGTCTTATGATCCCATTTGCAGAGAGCTCTCACTTGGTTATAGATAATA
CATAGTTACCAATGATGAAGCAATATAAACCCCAATTTCTAATTTGTAAATGAAGATAA
TAAACTACTTGCTGCATAGAGTTGCTGGGAAGATTAAATAAGTCCATATAGATGTAAAG
TGCTTAAACTATGCCAGACCTATGGTAAGTGACAAGAGTTGTTATTGGGATTTTAAAA
[T,-]
TATTATTATTATTATTATTATTATTGAGACAGAGTCTCGCTCTGTCTCCAGGCTGGAG
TGCAGTGGCGTGATCTCGGCTCACTGCAAGCTCCGCTCCAGGTTACGCCATTCTCTT
GCCTCAGCTCCCGAGTAGCTGGGACTACAGGCGCCGCACTACACCCGGCTAATGTTT
TGTATTTTATAGTACAGACAGGGTTTACCCTGTTATCCAGGATGGTCTCGATCTCCTGA
CCTCATGATCCACCCGCTTGTCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACCGC

6783 TGCGAGCGGCAGGTTTGCCTGATAATTCTTCTTGTCATAGTAGAGGCGGGATGTGGTAA
CAGAGACCAAGACTGTGGAGTTGGTGATTGTGGCTGATCACTCGGAGGTGAGCCTGCTGG
CCCCTGCACATCCTCCTCCCCCTGCACTGCCCTGCCGCTTTCATGTACCTCTCTTGGC
CTACAGGCCAGAAATACCGGACTTCCAGCACCTGCTAAACCGCACACTGGAAGTGGCC
CTCTTGCTGGACACAGTGAGTGCTGGACAGGGCAACCCCCACCCAGGCCCTGACCATG
[G,A]
CAACCCCTCTTCTGAGCCCCAGCTGTCTTTCAGTTCTTCCGGCCCCCTGAATGTACGAGTG
GCACTAGTGGGCTGGAGGCTGGACCCAGCGTGACCTGGTGGAGATCAGCCCAAACCA
GCTGTACCCCTCGAAAACTTCTCCTCACTGGCGCAGGGCACATTTGCTGCCTCGATTGCCC
CATGACAGTGCCAGCTGGTGACGTAAGGGCCCCAGACTCAGCCAGAGAGGCCAGTCTGT
TCCTGGCCAAATTCACACCCCTTCAGCACCTACCTCAGCCCTGAAGCTCTGACCACCG

7514 TATTTCCAGGTCTCCTCCTCATTCCCAATTCAGTTCTTCCCAAGTGTGGTGGCATTATG
CACTGAAACCCCCCTATAAAGTTGCCCAACCCCAAGCTACAGGTATAGAGGGTGGAGGT
ACGTGATGTGGCTTTGCTATCAGGGAGCCCTCGTTATGGCCAGCTAGTCACAGTGAC
ACAGTCATCCCCTGTGCACTCTTCCCATTTCTTAGAGGAGGGTAGGAGGCAGCTAAGGCC
CAAAGAACAGAGGTGATCTCCCTCCAGTGAGGGAGGGGGACAGAGCTGAGCTAGAACCCA
[A,C]
GTTTCTGCCATCCAGGCTGGGTTCTCCTACTTTAGAAGCAATTCAGGAGGGAAGCAGTG
CCTGCTGAGTGCCACGAGGTGACAGTGAGGGAACAGGAGCAGAGAGGGTGGTCTGGG
CATTGTGGTGGAGGCAGGCTGGGACTGGACCTACAGTACCCCTCCCAATGACAGGACCA

FIGURE 3J

CTCCACCAGCATCCTGGGAGTCGCCTCCTCCATAGCCCATGAGTTGGGCCACAGCCTGGG
CCTGGACCATGATTTGCCTGGGAATAGCTGCCCTGTCCAGGTCCAGCCCCAGCCAAGAC

15505

CGCCAGAAGACGCTGCCGCCTCAGACGAGGGCGGGCTGTGTGGGGCGGGAGTACCAGAAA
GGGTCGGCGTGTGTCCCCGGGATGCTCGCAGCTTCCCTCTGCCCAGACTGGGGTGGCTTT
CGGCGCAATCTGTCAAGCTGTTGGACCTGCCGTCCCCACTCTGACCATTGGCTGGGAAAA
GTGGATCTGGCTGATGCTCCAGAGCCCAGGAGCCAGGGCGGAGCGGGGCGGCGGCTGCT
CCCACGATCCCAAGGCCGCGCACCTGCCTCCTCCCCCTCCGCCGCCGCACTTGAGGGAT
[C,T]
GGGAACAAAGGTGCTTTGTACAGGCCGCAACCACCTCATTACTTCGTCTTAGGGACTGGG
GCCGCGTGGGCCCCCAGCCCGGAACGAAGGTGTGGAGCGGCAAGGGACAGACGCCAATCT
TAAAGTGAGCATCTAGCGCGCCACCTAAGGCTCTTTAGGGAAGGTGGTCCCAGAGCTGTG
TTGTCCTTCCGCTTGCACTGTCCCTAGATGTGCAAAGAAAACGGGGCAGTGCATGAAGG
TGTTGGACAGGCTTCATGGATCCTCGCCCGCGCTCACTTTCCTTATCTGGGCAAAGG

16123

AAATCTTCCAACTTCTAATAAGGCAGTCTACCCTGCACTAAAGCAGACACGAAAGAGAT
GACCTCCCTAAAAATACTGCTGTTGGAATACGTCTTCTTCCCGCCCCCTCGCAGTGCG
GTGCAGCCTCAGTGGAAGCTTTGGCGAACCTGGCGCGCGCTGCGGTGCACAGAGGGTTAA
CTGGAGTTGGCGCTGGGTGGAGAGGAGGAGACGCGCTCCCATTTGGCGGAAAGTTATTCAG
GGGCGGGGTGAGTGAATCTCCGTACCCCACTCCCTTTCCGCAACTTCCCTCTTCACTTT
[A,G]
TACCTTTCTCTCCTCGACTGTGAAGCGGGCCGGGACCTGCCAGGCCAGACCAAACCGGAC
CTCGGGGGCGATGCGGCTGCTGCCCCTGCTGCGGACTGTCTATGGGCCGCGTCTCGGC
TCCCCTCTGCGCGGGGGCTCCAGCCTCCGCCACGTAGTCTACTGGAATCCAGTAACCCC
AGGTAGCCGGGCGAACCAGGGCGAGCGCACAGCCAAGTCTGCGCGCTCCCGGGCTTTGCG
CGCGCCCGCACCCGCTCTTTGCGCGGCGCGCCTGAGCCTGGCCGCGCGCGGGGCTCC

Chromosome map:
Chromosome # 1